

APPLICATION
FOR
UNITED STATES LETTERS PATENT

TITLE: ENHANCING THE INTELLIGIBILITY OF RECEIVED
SPEECH IN A NOISY ENVIRONMENT

APPLICANT: ADORAM ERELL

CERTIFICATE OF MAILING BY EXPRESS MAIL

Express Mail Label No. EL688323013US

I hereby certify under 37 CFR §1.10 that this correspondence is being deposited with the United States Postal Service as Express Mail Post Office to Addressee with sufficient postage on the date indicated below and is addressed to the Commissioner for Patents, Washington, D.C. 20231.

1/30/2002

Date of Deposit

Signature

Gabriel Lewis

Typed or Printed Name of Person Signing Certificate

ENHANCING THE INTELLIGIBILITY OF RECEIVED SPEECH IN A NOISY ENVIRONMENT

TECHNICAL FIELD

This invention relates to audio signal processing for speech communication.

BACKGROUND

The increasing demand for constant personal communication has resulted in the availability of phones in just about any location imaginable. Wireless or mobile phones have enabled individuals to communicate while roaming in a variety of dynamic environments, such as airplanes, cars, restaurants, and other public and private places. Furthermore, wireline phones, such as public pay phones, can be found in a variety of different environments, such as airports, train stations, stores, and gas stations, among others. These and other such environments are associated with considerable ambient or background noise which makes it difficult to clearly transmit and receive intelligible speech at an audible level.

As a result, the individuals on a phone conversation may often have to repeat themselves, which is inconvenient, time-consuming, inefficient and costly. Alternatively, they may shout or raise their voices to be heard over the noise, which compromises the privacy of the conversation. A person in a noisy environment may also increase the volume of the phone in

order to better hear the person who is speaking on the other end.

Manually adjusting the volume level in response to loud background noise is tedious for both the individual at the speaking end and the individual at the listening end. Furthermore, manually increasing volume in response to background noise is undesirable since the volume must be later manually decreased to avoid acutely loud reception when the background noise dies down.

DESCRIPTION OF DRAWINGS

Figure 1 illustrates a communication system used for transmitting an audio signal from a far-end to a near-end.

Figure 2 illustrates an audio path using a receive side automatic volume control according to one embodiment of the present invention.

Figure 3 illustrates a process for determining the automatic volume control gain according to one embodiment of the present invention.

Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

Referring to the example in Figure 1, a far-end device 102 detects far-end sound 105 that can include speech. The

sound 105 is converted to a signal 106, the far-end signal, which is transmitted to the near-end device 101, for example, by modulating a radio frequency signal, interfacing with a network such as the Internet, or sending a signal on a waveguide. The transmission of the signal 106 can also include combinations of known signal transmission modes, such as those that use electric, optical, microwave, infrared, and radio signals, and any number of intermediaries, such as switches, computer servers, and satellites.

The near-end device 101 reproduces the far-end sound 105. The near-end device 101 also detects near-end sound that can include ambient noise 103. The near-end device 101 processes the signal 106 in response to the ambient noise 103 in order to render the far-end sound 105 more human-interpretable to a user of the near-end device 101.

In the example depicted in Figure 1, the near-end device 101 is a handheld telephone that receives the far-end signal 106 from the far-end device 102 which is a telephone at a remote location.

An audio path using automatic volume control (AVC) is illustrated in Figure 2. In the audio path, the near-end device 101 uses a microphone 205 to detect sound and ambient noise on the near-end. An analog signal for the near-end sound can be converted into a digital signal by a processor, CODEC 210. The

digital signal is then sent through an echo canceller 215 to remove any echo from the signal. The signal is then processed by a transmission side gain control 220 to determine if any gain is needed by the signal. The resulting signal is evaluated by a transmit voice activity detector (VAD) 225, then sent through a noise suppressor 230 prior to being encoded by an encoder 235 for transmission (TX). The transmission VAD 225 supplies a signal to a receive signal automatic volume control (RX-AVC) module 260 indicating the noise level at the transmission end.

The receive path receives the transmitted signal and decodes the received signal in the decoder 240. The decoded signal is then evaluated by the receive VAD 245 and processed by the receive enhanced noise suppressor 250 and an equalizer 255. The RX-AVC module 260 then automatically controls the received speech level and dynamic range. The AVC module 260 is a collection of three functions. The functions include an automatic gain control (AGC), an automatic volume increase and dynamic range compression (DRC) as a function of the transmit noise level, and a DRC for the speakerphone. After being processed by the AVC module 260, the signal is processed by a receive gain controller 265, and then converted to an analog signal by the CODEC 210. The analog signal is then rendered as sound by a speaker 270.

The total gain of the AVC module 260 is determined by summing three gains: the AGC, the DRC gain weighted by a maximum between a weight proportional to the transmit noise level and the DRC gain factor, and an additional gain whose value is weighted by the transmit noise level. Figure 3 illustrates a process for determining the total gain of the AVC module 260 according to one embodiment of the present invention.

Figure 3 illustrates a process 300 to determine the total AVC gain according to one embodiment of the present invention. The total AVC gain is the sum of three gains, the AGC gain, an additional constant gain weighted by the transmission noise level, and a weighted DRC gain. The process 300 begins in block 305 where the maximum sample energy of the decoded far-end speech over a sub frame is computed. The sub frame may consist of several samples, and may have a default number of samples such as 16. The number of samples in the sub frame represents a compromise between a desired fine time resolution for the gain update decisions (smaller sub frame) versus computation resources (larger sub frame).. In block 305, the process 300 also computes the RMS energy level over the sub frame.

Proceeding to block 310, the process 300 computes the signal envelope level and speech RMS. The minimum peak threshold and the RX-VAD detected speech are received as inputs

to the computation of the signal envelope level. The envelope is then updated if the RX-VAD detected speech and the local peak threshold are above a predetermined threshold level. The threshold level may also be dynamic. The pseudo code for updating the envelope is as follows:

```

if (RX-VAD && local_peak > MIN_ENVELOPE)
{
    if (local_peak > envelope)
        envelope = min(local_peak, envelope +
INCREMENT_PEAK);
    else
        envelope -= DECREMENT_PEAK'

    if (rms>long_term_rms)
        long_term_rms +=INCREMENT_RMS;
    else
        long_term_rms -=INCREMENT_RMS;
}

```

MIN_ENVELOPE is provided to ensure that signals whose envelopes are smaller than this threshold do not affect the AGC.

Proceeding to block 315, the process 300 computes the AGC gain. The basic relation for the AGC gain is defined as $agc_gain = (LEVEL1 - envelope)$. LEVEL1 is the target level for the receive signal envelope level. The default value is 3 dB

below the clipping level. The basic relation includes two limitations. The first limitation is that the $\text{agc_gain} = \min(\text{MAX_AGC_GAIN}, \text{agc_gain})$, or that the AGC gain cannot exceed a maximum value. MAX_AGC_GAIN is the upper limit on the AGC gain, with a default value of 21 dB. The second limitation is $\text{agc_gain} = \min(\text{MAX_RMS} - \text{long_term_rms}, \text{agc_gain})$, or that the AGC gain is the lesser of the calculated AGC gain and the maximum RMS minus the long term RMS. MAX_RMS is provided so the AGC gain is limited to the signal RMS after amplification, with a default of -15 dB below clipping. The AGC gain is then included as one input to the summer 350.

Proceeding to block 335, the process 300 computes the DRC gain. The instantaneous level and the envelope level are received as inputs to the computation of the DRC gain. The DRC gain computation is given the noise dependent and receive signal independent parameters:

$\text{drc_gain} = \text{MAX_DRC_GAIN} * \max(\text{drc_gain_factor}, \text{noise factor})$.

The MAX_DRC_GAIN is the upper limit on the DRC gain. A default value may be set to the maximum allowable 21 dB. Decreasing this parameter decreases the effect of the DRC. The drc_gain_factor controls the amount of DRC.

For each sub-frame, the DRC gain of the frame is calculated as a function of the local peak and the envelope

level. The resulting raw DRC gain is supplied to the multiplier 345.

Proceeding to block 320, the near-end microphone signal is received and the transmission noise level is estimated. An input from the TX VAD is used to estimate the transmission noise level.

Proceeding to block 325, the transmission noise level is used to compute a transmission noise factor, which in one embodiment may be a number between 0 and 1. The noise factor is a function of the noise level. For example, with no ambient transmission noise, the transmission noise factor will be 0. As the transmission noise increases, the noise factor increases until a maximum transmission noise level is reached. At the maximum transmission noise level, the noise factor is 1. The noise factor is provided as an input to the multiplier 330 and the MAX block 340.

Proceeding to multiplier 330, a constant additional gain is multiplied by the noise factor to obtain an additional constant gain whose value is weighted by the transmission noise level. The weighted constant gain is then provided as a second input to the adder 350.

Proceeding to block 340, a maximum is determined between the noise factor and the DRC Gain Factor. The larger of the two factors is then supplied to the multiplier 345 to weight

the computed DRC gain. The resultant weighted DRC gain is then supplied as a third input to the adder 350.

The adder 350 combines the AGC gain, the weighted DRC gain, and the weighted additional constant gain to determine the total AVC gain. The total AVC gain may then be smoothed over a few frames using signed exponential smoothing using the following pseudo code:

```

    if (target_gain > smooth_gain_db)
        alfa = ALFA_UP;
    else
        alfa = ALFA_DOWN;
    smooth_gain_db = smooth_gain_db * (1-alfa) + alfa *
target_gain;

```

The result is a noise weighted and smoothed total AVC gain.

The techniques may be implemented in hardware, software, or a combination of the two in order to analyze digital or analog signals. The techniques described here are also not limited to telephones, or the exemplary configuration described above; they may find applicability in any computing or processing environment for communications. For example, desktop computers linked to a computer network can be used to exchange sound communications that include human speech and ambient noise. Typically, each device may include a sound input device,

such as a microphone, and a sound output device, such as a loudspeaker.

A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

10559-335001-P9850